

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 19, 2002, 07:53:17 ; Search time 1603 seconds  
(without alignments)  
3856.273 Million cell updates/sec

Title: US-09-807-459-2  
Perfect score: 2359  
Sequence: 1 MAPSDVGDVTKTLTAASES:.....DPSKALIRKVTSTADNLEK 458

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 13736207 segs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ -p2n model -DEV-xlh  
-Q/cgn\_1/OSPT0.us09807459/unat\_18102002.141112\_28595/app-query.fasta.1.647  
-DB-EST -QFMT-fastap -SUFFIX-rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS-bits -FASTAP -END=-1 -MATRIX-biosum62 -TRANS-human40.cdi -LIST=45  
-DOCLALIGN=200 -THR SCORE-pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE-LOCAL  
-OUTFMT-pto -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=us09807459.ecgn.1.1\_753-etunal.18102002.141112\_28595 -ICPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAP -LARGEQUERY -NEG-SCORES=0 -WAIT -LONGLOG -DEV-TIMEOUT=120  
-WARN-TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:.\*  
2: em\_esthum:.\*  
3: em\_estln:.\*  
4: em\_estmu:.\*  
5: em\_estov:.\*  
6: em\_estpl:.\*  
7: em\_estro:.\*  
8: em\_hlc:.\*  
9: gb\_est1:.\*  
10: gb\_est2:.\*  
11: gb\_hlc:.\*  
12: gb\_gss:.\*  
13: em\_gss\_hum:.\*  
14: em\_gss\_inv:.\*  
15: em\_gss\_pln:.\*  
16: em\_gss\_vrt:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	ID
1	131	5.6	529 12 A2167872
2	123	5.2	564 10 BM275292

C	3	120	5.1	952	12	A2209422	A2209422 SP_0106_A
C	4	118	5.0	856	12	A2199582	A2199582 SP_1039_B
C	5	117.5	5.0	1035	10	BM416596	BM416596 Op21695 M
C	6	117	5.0	504	10	BM274689	BM274689 PFEStroa7
C	7	117	5.0	537	10	BM274676	BM274676 PFEStroa7
C	8	115	4.9	366	12	A2140870	A2140870 SP_0001_B
C	9	113	4.8	599	12	A2199581	A2199581 SP_1039_B
C	10	113	4.8	665	12	A2400717	A2400717 IM0167E11
C	11	112.5	4.8	897	12	CNS02TD6	AL213027 Tetradon
C	12	112	4.7	624	9	AA571652	AA571652 vm09c03_r
C	13	111	4.7	794	12	BH098592	BH098592 RPT1-24-2
C	14	110.5	4.7	698	12	A2766807	A2766807 IM0564C11
C	15	109	4.6	324	12	A2152840	A2152840 SP_0034_B
C	16	109	4.6	642	12	A2988471	A2988471 2M0271021
C	17	109	4.6	700	9	BB230024	BB230024 BB230024
C	18	109	4.6	943	11	BC019933	BC019933 Mus muscu
C	19	108.5	4.6	721	12	A2814810	A2814810 2M0082P07
C	20	108	4.6	996	12	CNS06H10	AL439038 T3 end of
C	21	108	4.6	1154	12	CNS07C18	AL438594 T3 end of
C	22	107.5	4.6	642	12	A2431579	A2431579 IM0216006
C	23	106	4.5	640	12	A2733090	A2733090 RPT1-24-1
C	24	105	4.5	478	12	A2462925	A2462925 IM0271A02
C	25	104.5	4.4	532	10	BM273998	BM273998 PFEStroa6
C	26	104	4.4	596	12	A2944493	A2944493 2M0205C06
C	27	104	4.4	644	10	BM168451	BM168451 EST570974
C	28	103.5	4.4	715	12	BH126086	BH126086 RPT1-24-2
C	29	103	4.4	717	10	BG523384	BG523384 31-25 Ste
C	30	103	4.4	791	12	BH047254	BH047254 RPT1-24-2
C	31	102.5	4.3	439	12	A2946473	A2946473 2M0208B13
C	32	102.5	4.3	503	12	A2344512	A2344512 IM0078M05
C	33	102.5	4.3	547	12	A2696548	A2696548 RPT1-23-2
C	34	102	4.3	869	12	A2679969	A2679969 ENTE206TR
C	35	101.5	4.3	683	12	BH070091	BH070091 RPT1-24-2
C	36	101.5	4.3	830	12	A2667704	A2667704 ENTKT09TF
C	37	101	4.3	605	12	A2891746	A2891746 RPT1-24-2
C	38	101	4.3	614	12	A2415896	A2415896 IM0190A17
C	39	101	4.3	622	12	A2625698	A2625698 IM0465N20
C	40	101	4.3	700	12	BH313240	BH313240 CH230-101
C	41	100.5	4.3	613	12	A2342112	A2342112 IM0075F05
C	42	100.5	4.3	651	12	BH352781	BH352781 CH230-175
C	43	100.5	4.3	658	12	A2421595	A2421595 IM0199K23
C	44	100.5	4.3	710	12	BH278149	BH278149 CH230-48P
C	45	100.5	4.3	739	12	BH036938	BH036938 RPT1-24-3

## ALIGNMENTS

RESULT 1	A2167872	529 bp	DNA	linear	GSS 29-AUG-2000
LOCUS	SP_0103_A2_E10_T7A	Strongylocentrotus purpuratus			, sperm genomic BAC library Strongylocentrotus purpuratus genomic
DEFINITION	clone Plate=103 Col=20 Row=I,	DNA sequence.			
ACCESSION	A2167872				
VERSION	A2167872.1	GI:8338240			
KEYWORDS	GSS.				
SOURCE	Strongylocentrotus purpuratus.				
ORGANISM	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae; Strongylocentrotus.				
REFERENCE	1 (bases 1 to 529)				
AUTHORS	Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Bondi, T.R., Swartzell, S., Wallace, J.C., Pousta, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.				
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources				
JOURNAL	Proc. Natl. Acad. Sci. U. S. A.	97 (17)			9514-9518 (2000)
MEDLINE	20402566				
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L				
	Division of Biology 156-29				
	California Institute of Technology				



DB 184 CCAACAAATGACCTTATAATGACCAATAATGACCAACAAATGACCAATAATGAC 243  
 QY 366 GlllleArgAspProSerlySAlaleuileArglyValSerThrlyAlaGluAspLeu 365  
 DB 244 CCAATTAATGACCAACCAATGACCAATAA-----AATGACCCA 282  
 QY 386 PheGluAsnlyIleGlyInglyThrValAspPheIleAsnngluileArgAspPro 405  
 DB 283 ATTAATGACCAATAATGTCACCAATGACCAATAATGACCAATAATGACCA 342  
 QY 406 SerlySAlaleuileArglyValTyrrhGluAlaAspAspLeupheGluAsnlyIle 425  
 DB 343 ATTAATGTCACCAATAAT-----GACCAATAATGACCAATAATGACCAATA 393  
 QY 426 GlyInglyThrValAspPheIleAsnlyGluileArgAspPro 440  
 DB 394 AATGACCAACAAATGACCAATAATGACCAATAATGACCA 438

RESULT 3  
 AZ209422/c 952 bp DNA linear GSS 31-AUG-2000  
 LOCUS SP\_0106\_A2\_A12-SP6E Strongylocentrotus purpuratus, purple sea  
 DEFINITION urchin, sperm genomic BAC library Strongylocentrotus purpuratus  
 genomic clone Plate-106 Col=24 Row=A, DNA sequence.  
 AZ209422  
 VERSION AZ209422.1 GI:8423111

KEYWORDS GSS.  
 SOURCE Strongylocentrotus purpuratus.  
 ORGANISM Strongylocentrotus purpuratus.  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 Echinoidea; Euechinoidea; Echinoidea; Echinoidea;  
 Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 952)  
 Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,  
 Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray,  
 G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and  
 Hood, L.  
 A sea urchin genome project: Sequence scan, virtual map, and  
 additional resources  
 Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)  
 20402566  
 JOURNAL Contact: Cameron, RA, Davidson, EH, Hood, L  
 MEDLINE Division of Biology 156-29  
 COMMENT California Institute of Technology  
 Pasadena California 91125, USA  
 Tel: (626) 395-8421  
 Fax: (626) 793-3047  
 Email: acameron@caltech.edu  
 Plate: 106 Row: A Column: 24  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence stop: 952.

FEATURES  
 source  
 1..952  
 Location/Qualifiers  
 /organism="Strongylocentrotus purpuratus"  
 /db\_xref="taxon:7668"  
 /clone="Plate-106 Col=24 Row=A"  
 /clone\_lib="Strongylocentrotus purpuratus, purple sea  
 urchin, sperm genomic BAC library"  
 /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli  
 DH10B"

BASE COUNT 145 a 265 c 270 g 272 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.00381 Length: 952  
 Score: 120.00 Matches: 38  
 Percent Similarity: 42.57% Conservative: 25  
 Best Local Similarity: 25.68% Mismatches: 69  
 Query Match: 5.09% Indels: 16  
 DB: 12 Gaps: 4

US-09-807-459-2 (1-458) x AZ209422 (1-952)

QY 294 AsnlyGluileArglyPurProSerlySAlaleuileGlyValSerThrlyAspThrlyS 313  
 DB 523 AACAGGCCATCGAGACCCCAAGTCGACCGCTGCTGCGATGGACACCGCA 464  
 QY 314 AspLeupheGluAsnlyIleGlyInglyThrValAspPheIleAsnngluileArg 333  
 DB 463 GAATCATGAGAGAACCGGATTTCTTCGGGGAC-----AACCAAGCATCGAG 416  
 QY 334 AspProSerlySAlaleuileArglyValSerAsnAspAlaleuileArglyVal 353  
 DB 415 ACTCAAGTCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 356  
 QY 354 LysIleGlyInglyThrValAspPheIleAsnngluileArgAspProSerlySAla 373  
 DB 355 CGGTATTCCTTTGGGGAT-----AACCAAGCATCGAGACTCCACCC 308  
 QY 374 LeuileArglyValSerThrlyAlaGluAspLeupheGluAsnlyIleGlyIngly 393  
 DB 307 TCGCCCGTGTCTTCGACGGGACAGCCGAGATCGATGAGAACCGGATTTCTTCGGGG 248  
 QY 394 ThrValAspPheIleAsnngluileArgAspProSerlySAlaleuileArglyVal 413  
 DB 247 GCC-----AACCAAGCATCGAGACTCCACCGCTGCTGCTGCTGCTG 200  
 QY 414 TyrrhGluAlaAspAspLeupheGluAsnlyIleGlyInglyThrValAspPheIle 433  
 DB 199 GCATGGGACGCGGATTCGATGAGAACCGGATTCCTTTGGGGAT----- 152  
 QY 434 AsnlyGluileArgAspProSer 441  
 DB 151 AACCAAGCATCGAGACTCCACGT 128

RESULT 4  
 AZ199582/c 856 bp DNA linear GSS 31-AUG-2000  
 LOCUS SP\_1039\_B2\_E12-77A Strongylocentrotus purpuratus, purple sea urchin  
 DEFINITION , sperm genomic BAC library Strongylocentrotus purpuratus genomic  
 clone Plate-1039 Col=24 Row=J, DNA sequence.  
 AZ199582  
 VERSION AZ199582.1 GI:8394390

KEYWORDS GSS.  
 SOURCE Strongylocentrotus purpuratus.  
 ORGANISM Strongylocentrotus purpuratus.  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 Echinoidea; Euechinoidea; Echinoidea; Echinoidea;  
 Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 856)  
 Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,  
 Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray,  
 G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and  
 Hood, L.  
 A sea urchin genome project: Sequence scan, virtual map, and  
 additional resources  
 Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)  
 20402566  
 JOURNAL Contact: Cameron, RA, Davidson, EH, Hood, L  
 MEDLINE Division of Biology 156-29  
 COMMENT California Institute of Technology  
 Pasadena California 91125, USA  
 Tel: (626) 395-8421  
 Fax: (626) 793-3047  
 Email: acameron@caltech.edu  
 Plate: 1039 Row: J Column: 24  
 Seq primer: 77  
 Class: BAC ends  
 High quality sequence stop: 856.

FEATURES  
 source  
 1..856  
 Location/Qualifiers  
 /organism="Strongylocentrotus purpuratus"  
 /db\_xref="taxon:7668"  
 /clone="Plate-1039 Col=24 Row=J"



DEFINITION PFESToa73e05.y1 Plasmodium falciparum 3D7 gametocyte cDNA library  
Plasmodium falciparum 3D7 cDNA 5' similar to TR:097255 097255  
MALP3.1 PROTEIN.; mRNA sequence.  
ACCESSION BM274689  
VERSION BM274689.1 GI:17968000  
KEYWORDS EST.  
SOURCE Plasmodium falciparum 3D7.  
ORGANISM Plasmodium falciparum 3D7.  
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
AUTHORS 1 (bases 1 to 504)  
Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,  
Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,  
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,  
Tsagarisshvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,  
Maguire,L., Ritchey,J., Wadkins,J., Kennedy,S., Levins,D.,  
Waterston,R., Wilson,R. and Sibley,D.  
Mashu Plasmodium EST Project  
Unpublished (2001)  
Contact: L. David Sibley  
Mashu Plasmodium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library was constructed by R. Haywood. DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: -40UP from Gibco  
High quality sequence stop: 416.  
Location/Qualifiers  
1..504  
/organism="Plasmodium falciparum 3D7"  
/db\_xref="taxon:36329"  
/clone\_lib="Plasmodium falciparum 3D7 gametocyte cDNA  
library"  
/dev\_stage="gametocyte (stage III-V)"  
/lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2:  
XhoI; The library was constructed by R Haywood. cDNAs were  
synthesized from gametocyte poly(A)+ RNA by oligo d(T)  
priming, size-selected and directionally cloned into the  
EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR  
lambda vector (Stratagene). The primary library was mass  
excised as phagemid using the ExAssist helper phage  
(Stratagene). Clones were mass excised using the ExAssist  
helper phage (Stratagene), the phagemids were precipitated  
with PEG 8000 and extracted with phenol/chloroform.  
Phagemid DNA was electroporated into DH10B cells. Clone  
Availability: David Sibley, Washington University."

BASE COUNT 253 a 130 c 42 g 78 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 0.00316 length: 504  
Score: 117.00 matches: 41  
Percent Similarity: 38.71% conservative: 19  
Best Local Similarity: 26.45% mismatches: 81  
Query Match: 4.96% indels: 14  
DB: 10 gaps: 4

US-09-807-459-2 (1-458) x BM274689 (1-504)

OY 287 ILEProthLysLysPheAsnLysGluIleArgLuproSerLysAlaLeuLysGlu 306  
DB 13 GTCCCAACAATGACCAATTAATGACCAATTAATGACCA-----ATAATGTC 63  
OY 307 LysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp 326  
DB 64 CCAATTAATGACCAATTAATGACCAATTAATGACCAATTAATGAC 123  
OY 327 PhePheAsnLysGluIleArgAspPro---SerLysAlaLeuLysGluLysValSerAsn 345

DB 124 CCAATTAATGACCAATTAATGACCAATTAATGACCAATTAATGACCAATTAATGAC 183  
OY 346 AspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheLeaAsn 365  
DB 184 CCAACAAATGACCAATTAATGACCAATTAATGACCAATTAATGACCAATTAATGAC 243  
OY 366 GluIleArgAspProSerLysAlaLeuIleArgLysValSerThrLysAlaGluAspLeu 385  
DB 244 CCAATTAATGACCAATTAATGACCAATTAATGACCAATTAATGACCAATTAATGAC 282  
OY 386 PheGluAsnLysIleGlyGlnGlyThrValAspPheLeaAsnGluIleArgAspPro 405  
DB 283 ATTAATGACCAATTAATGACCAATTAATGACCAATTAATGACCAATTAATGACCA 342  
OY 406 SerLysAlaLeuIleArgLysValThrGluAlaAspAspLeuGluAsnLysIle 425  
DB 343 ATTAATGTCCTCAATTAAT-----GACCAATTAATGACCAATTAATGACCAATTA 393  
OY 426 GlyGlnGlyThrValAspPheLeaAsnLysGluIleArgAspPro 440  
DB 394 AATGACCAACAATTAATGACCAATTAATGACCAATTAATGACCA 438

RESULT 7  
BM274676  
LOCUS 537 bp mRNA linear EST 20-DEC-2001  
DEFINITION PFESToa73e03.y1 Plasmodium falciparum 3D7 gametocyte cDNA library  
Plasmodium falciparum 3D7 cDNA 5' similar to TR:097255 097255  
MALP3.1 PROTEIN.; mRNA sequence.  
ACCESSION BM274676  
VERSION BM274676.1 GI:17967987  
KEYWORDS EST.  
SOURCE Plasmodium falciparum 3D7.  
ORGANISM Plasmodium falciparum 3D7.  
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
AUTHORS 1 (bases 1 to 537)  
Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,  
Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,  
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,  
Tsagarisshvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,  
Maguire,L., Ritchey,J., Wadkins,J., Kennedy,S., Levins,D.,  
Waterston,R., Wilson,R. and Sibley,D.  
Mashu Plasmodium EST Project  
Unpublished (2001)  
Contact: L. David Sibley  
Mashu Plasmodium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library was constructed by R. Haywood. DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: -40UP from Gibco  
High quality sequence stop: 423.  
Location/Qualifiers  
1..537  
/organism="Plasmodium falciparum 3D7"  
/db\_xref="taxon:36329"  
/clone\_lib="Plasmodium falciparum 3D7 gametocyte cDNA  
library"  
/dev\_stage="gametocyte (stage III-V)"  
/lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2:  
XhoI; The library was constructed by R Haywood. cDNAs were  
synthesized from gametocyte poly(A)+ RNA by oligo d(T)  
priming, size-selected and directionally cloned into the  
EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR  
lambda vector (Stratagene). The primary library was mass  
excised as phagemid using the ExAssist helper phage  
(Stratagene). Clones were mass excised using the ExAssist

helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells. Clone Availability: David Sibley, Washington University." 128 46 82 +

BASE COUNT	271 a	138 c	46 g	82 t
ORIGIN				

**Alignment Scores:**

Pred. No.:	0.00349	Length:	537
Score:	117.00	Matches:	41
Percent Similarity:	38.71%	Conservative:	19
Best local Similarity:	26.45%	Mismatches:	81
Query Match:	4.96%	Indels:	14
DB:	10	Gaps:	4

US-09-807-459-2 (1-458) x BM274676 (1-537)

OY	287	IlePprrrrLysLysPhePheasnLysGluIleargGluProSerLyAlaleuLysGlu	306
Db	13	GTCACCAACAATGACCACCAATAATAGCCCATTAATGACCCA-----ATAAATGC	63
OY	307	LysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp	326
Db	64	CCAAATAAATGACCCCAATAAATAGCCCAATAAATAGCCCAATAAATAGC	123
OY	327	PhePheAsnLysGluIleArgAspro---SerLysAlaleuLysGluLysValSerSrn	345
Db	124	CCAAATAAATGACCCCAATAAATAGCCCAATAAATAGCCCAATTAATATAC	183
OY	346	AspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnSn	365
Db	184	CCAACAAATGACCAAAATAATAGCCCAATTAATAGCCCAACAAATGACCAATAAATGAC	243
OY	366	GluIleArgAsproSerLysAlaleuIleArgLysValSerThrGlyAlaGluAspLeu	385
Db	244	CCAAATAAATGACCCCAACAATGACCAATA-----AATGACCA	282
OY	386	PheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspro	405
Db	283	ATAAATGACCCCAATAAATGTCCCAACAATAATGACCAATAATGACCCCAATAATGACCA	342
OY	406	SerLysAlaleuIleArgLysValTyThrGluAlaAspAspLeuPheGluAsnLysIle	425
Db	343	ATAAATGTCCCAATTAAT-----GACCCAAATAAATGACCCCAATAATGACCAATA	393
OY	426	GlyGlnGlyThrValAspPheIleAsnLysGluIleArgAspro	440
Db	384	AATGACCCCAACAATTAATTAATGACCCCAATAAATGACCCA	438

RESULT 8	LOCUS	DEFINITION
A2140870	366 bp	DNA linear
A2140870	SP_0001_B2_C01_SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library strongylocentrotus purpuratus genomic clone Plate1 C01-2 Row-N, DNA sequence.	GSS 28-AUG-2000

ACCESSION	A2140870
VERSION	A2140870.1
KEYWORDS	GSS.
SOURCE	Strongylocentrotus purpuratus.
ORGANISM	Strongylocentrotus purpuratus

REFERENCE	1 (bases 1 to 366)
AUTHORS	Cameron, R.A., Mahaltras, G., Rast, J.P., Martinez, P., Blondi, T.R.,

TITLE	JOURNAL	PAGE	DATE
A sea urchin genome project: Sequence scan, virtual map, and additional resources	Proc. Natl. Acad. Sci. U. S. A.	97 (17)	9514-9518 (2000)
20402366			
Contact: Cameron, RA, Davidson, EH, Hood, L			
COMMENT			

Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu  
Plate: 1 row: N column: 2  
Seq primer: Sp6  
Class: BAC ends  
High quality sequence stop: 366.  
Location/Qualifiers  
1..366  
FEATURES  
source

**source**

993.1.366

```
/db_xref="taxon:7668"
```

BASE COUNT ORIGIN	90 a	106 c	108 g	62 t
urcinin; sperm genomic BAC library				
/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli				
DH10B"				

Alignment Scores:	
Pred. No.:	0.00328
Score:	115.00
Percent Similarity:	44.62%
Best Local Similarity:	25.38%
Query Match:	4.67%
DB:	12
Length:	366
Matches:	33
Conservative:	25
Mismatches:	60
Indels:	12
Gaps:	1

US-09-807-459-2 (1-458) X AZ140870 (1-366)

QY	294	AsnLysGluLeuIea	ngLuproSerLysAla	LeuLysLulysValSerThrAspThrLys	313
Db	11	AACCAACCATCAG	ACTCCAAAGTCAC	AGCCTCGGCGTGTCTCTGGCATGGACACCGCA	70
QY	314	AspLeuPheGluAsn	LysLleGlyngIY	ThrValAspPhePheAsnLysGluLea	333
Db	71	GATTCGATGGA	GAACCCGATTTCTT	TGGGGAT-----AACCAAGCATCAAC	118
QY	334	AspProSerLysAla	LeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsn	353	
Db	119	ACTCCAAAGTCA	ACCCTCGCGCGTGTCT	GTGGACAGGACAGCAGCAATGATGAGAAC	178
QY	354	LysLleGlyngIY	ThrValAspPheLleAsnAsnLulLea	ngAspProSerLysAla	373
Db	179	CGGATATCTT	CGGGGAC-----AACCAAGCCATCAAA	CCCAAGTCCAGCC	226
QY	374	LeuIleA	ngLysValSerThrGlyAlaGluAspLeuPheGluAsnLysLleGlyngIY	393	
Db	227	TTCGGCCGTGT	CTCTGGCATGGAGAC	CGCGAGAACCATGGAGAACCATGTTCTT	CGGG
QY	394	ThrValAspPheLle	AsnAsnGluLleA	ngAspProSerLysAlaLeuLleA	ngLysVal
Db	287	GAT-----	AACCAAGCCATCGAGAC	CCCAAGTCCAGCCTCGCGTGTCTGTG	334
QY	414	TyrThrGluAla	AspAspLeuPheGluAsn	423	
Db	335	GCATGGACAG	CCGAGATCATGATGAGAGAC	364	

RESULT 9					
AZ199581					
LOCUS					
DEFINITION					
	AZ199581	599 bp	DNA	linear	GSS 31-AUG-2000
	SP_1039_B2_E12	SP66	Strongylocentrotus purpuratus	purpuratus	purple sea
	urchin	sperm	genomic BAC library	Strongylocentrotus purpuratus	
	genomic clone	Plate-1039	COL-24	Row-J	DNA sequence.

ACCESSION	A2199581
VERSION	A2199581.1
KEYWORDS	GI:6394385
SOURCE ORGANISM	GSS.
	Strongylocentrotus purpuratus.
	Strongylocentrotus purpuratus
	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinomorphia; Echinacea; Echinoida; Echinoidae; Echininoidea; Echinacea; Echinoida; Echinoidae;
	Echinozoa

REFERENCE	AUTHORS	TITLE	JOURNAL MEDLINE	COMMENT
1 (bases 1 to 599)	Strongylocentrotidae; Strongylocentrotus.			
	Cameron, R.A., Mahaias, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartwell, S., Wallace, J.C., Pouska, A.J., Livingston, B.T., Wray, G.A., Etensohn, C.A., Lehnach, H., Britten, R.J., Davidson, E.H. and Hood, L.	A sea urchin genome project: Sequence scan, virtual map, and additional resources	Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)	
		Contact: Cameron, Ra, Davidson, EH, Hood, L		
		Division of Biology 156-29		
		California Institute of Technology		
		Pasadena California 91125, USA		
		Tel: (626) 395-8421		
		Fax: (626) 793-3047		
		Email: acamerone@caltech.edu		
		Plate: 1039 row: J column: 24		
		Seq primer: SP6		
		Class: BAC ends		
		High quality sequence stop: 599.		
FEATURES		Location/Qualifiers		
source		1..599		
		/organism="Strongylocentrotus purpuratus"		
		/db_xref="taxon:7668"		
		/clone="Plate=1039 COL=24 Row=J"		
		/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"		
		/note="Organ: sperm; Vector: BAC63.6; BAC clones in E-Coli DH10B"		
BASE COUNT	154 a 168 c 167 g 110 t			
ORIGIN				
Alignment Scores:				
Pred. NO.:	0.0121	Length:	599	
Score:	113.00	Matches:	30	
Percent Similarity:	45.22%	Conservative:	22	
Best Local Similarity:	26.09%	Mismatches:	51	
Query Match:	4.79%	Indels:	12	
DB:	12	Gaps:	3	
US-09-807-459-2 (1-458) x A2199581 (1-599)				
OY	294 AsnLysGluIleAArgGluProSerLysAlaLeuLysGluLysValSerThrAspHrLys		313	
DB	11 AACCAAGCATTGACATCGAATCGAATCGATCGGCGGTCTCTGCGATGGACAGCGGA		70	
OY	314 AspleuphegIuAsnLysLleIleGlyncIythrValAspPheAsnLysGluIleArg		333	
DB	71 GATTGATGAGGAAACCGGATTCTTTGGGGGAT		118	
OY	334 AspProSerLysAlaLeuLysGluLysValSerAsnAlaLeuLysAspleuphegIuAsn		353	
DB	119 ACTCCAAAGTCAACCTCGCGCGTCTCTGCGAGGGGACAGCGGAATCGATGGAGAAC		178	
OY	354 LysLleGlyncIythrValAspPheIleAsnGluIleArgAspProSerLysAla		373	
DB	179 CGGATTCTTGGGGGAC		226	
OY	374 LeuIleArgLysValSerThrGlyAlaGluAspleuphegIuAsnLysLleGlyncIy		393	
DB	227 TCGGCGGTGTCCTGCGATGGAGACGCCGACAGAGAAACGATATCTTCGGGG		286	
OY	394 ThrValAspPheIleAsnGluIleArgAspProSerLysAla		408	
DB	287 GAT		319	
RESULT 10				
LOCUS	A2400717	665 bp	DNA	linear
DEFINITION	1M0161E1F Mouse 10kb plasmid U06C1M library Mus musculus genomic			
ACCESSION	A2400717			

VERSION	AZ400717.1	GI:10515791
KEYWORDS	GSS.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Mollusca; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A. and Wright,D., Weiss,R.	
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts	
JOURNAL	unpublished (2000)	
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0167 row: E column: 11 Seq primer: CGTTGTAAACGACGCGCAGT Class: plasmid ends High quality sequence stop: 665. Location/Qualifiers 1..665 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="U06C1M0167E11" /clone_lib="Mouse 10kb plasmid U06C1M library" /sex="Male" /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-" /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b1A7129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."	
BASE COUNT	301 a 132 c 122 g 110 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	0.0142	Length: 665
Score:	113.00	Matches: 58
Percent Similarity:	41.94%	Conservative: 46
Best local Similarity:	23.39%	Mismatches: 76
Query Match:	4.79%	Indels: 68
DB:	12	Gaps: 13
US-09-807-459-2 (1-458) x AZ400717 (1-665)		
QY 222 SerArgLeuGlnHisIleThrSerSerIyrIysAspIyrMetAspThrGlnIleProAla 241		
Db 21 GCAAACGTTAAGAACTTACTTACCAAGAAACCAAGGCCACTCACCATCAGACAAACCCAGC 80		
QY 242 LeuProlYsPheAlaIyAsrGpHeSerLeuMetValIaIglNargLeuAlaThrVal 261		

```

Db      81  ATTCCTCAACTAAGCAGACT-----
Qy      262  ALGLYTRValAspThrProTyrLysTyrIleMetLysLeuLysAsnPheMet 281
          |||||
Db      99  CCGGATACCCCAACACACACCC-----AAAAGCAACAGCTCGGATTAAAGAACTATCTC 152
          |||||
Qy      282  ValAsnArgValPheIleProThrLysLysPheAsn-----LysGIuIleArgGIu 299
          |||||
Db      153  ATGATGCTGTAGAGACTTTAAGAAAGGCAATTATTAACACTTAAGAAATACAGAG 212
          |||||
Qy      300  ProSer-----LysAlaLeuLysGIuLysValSerThrAspThrLysAspLeu 315
          |||||
Db      213  AACACTGCCAAAAGAGTAGAAGTCTTAAGAAATATAGAAAAACACACCAACAGGTG 272
          |||||
Qy      316  PheGIuLysLysIleGIuGIuLysThrValAspPhePheAsnLysGIuIleArgAspPro 335
          |||||
Db      273  ATGGAA-----TTTAACAAAACCATCCCAAGACCTA 302
          |||||
Qy      336  SerLysAlaLeuLys-----GIuLysValSerAsnAspAlaLys-----AspLeu 350
          |||||
Db      303  AAAAGGAGAGTAGAAATATATAGAAAAACGAAAGTCAACTGTGAGATAGAAAC 362
          |||||
Qy      351  PheGIuLysLysIleGIuGIuLysThrValAsp---PheIleAsnAsnGIuIleArgAsp 369
          |||||
Db      363  CTAGAAGATATA-----TCAGAAACCATAGATCAATCATCACTAACAGATACAGAG 416
          |||||
Qy      370  ProSerLysAlaLeuIleArgLysValSerThrGIuLysGIuAspLeuPheGIuAsn--- 388
          |||||
Db      417  ATGCAAGAGAGATC-----TCAGTGCAGAAAGTCCCAAGACAGACATG 461
          |||||
Qy      389  -----LysIleGIuGIuLysThrValAspPhe 397
          |||||
Db      462  GGCACACCAATCAAAAGAAATGCAAAATGAAAAAGATCCTAATCAAAATATCCAGGA 521
          |||||
Qy      398  IleAsnAsnGIuIleArgAspProSerLysAlaLeuIleArgLysValTyrThrGIuAla 417
          |||||
Db      522  GCCCAGGACACATGAGAGACCAAC-----CTACGATATATAGGATGAGAAAG 572
          |||||
Qy      418  AspAspLeuPheGIu-----AsnLysIleGIuGIuLysThr 429
          |||||
Db      573  AATGAAGATTTTCAAACTTAAGGGCCAGCAAAATCTTCAACAAATATATAGACGGGAA 632
          |||||
Qy      430  ValAspPheIleAsnLysGIuIle 437
          |||||
Db      633  TTCCCATACCTTAAGAAAGAGATG 656
          |||||

```

RESULT 11  
CNS02TD6 897 bp DNA linear GSS 15-MAY-2000  
LOCUS  
DEFINITION Tetradon nigroviridis genome survey sequence PUC-ori end of clone 16314 of library G from Tetradon nigroviridis, genomic survey sequence.  
ACCESSION AL213027.1 GI:7871846  
VERSION  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetradon nigroviridis.  
ORGANISM Tetradon nigroviridis.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.  
1 (bases 1 to 897)  
Roest-Crolius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.  
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 897)  
AUTHORS Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Winkler,P., Brotlier,P., Quetier,F., Saurin,W. and Weissenbach,J.

```

TITLE Human gene number estimate provided by genome wide analysis using
JOURNAL Tetradon nigroviridis DNA sequence
REFERENCE Unpublished
AUTHORS 3 (bases 1 to 897)
TITLE Genoscope.
JOURNAL Direct Submission
COMMENT Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetradon.
FEATURES
source
1. 897
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="16314"
/clone_11b="g"
/note="Genoscope sequence ID : COAG163DE07SP1-end :
PUC-ori"
BASE COUNT 301 a 179 c 244 g 161 t 12 others
ORIGIN
Alignment Scores:
Pred. No.: 0.0258 Length: 897
Score: 112.50 Matches: 77
Percent Similarity: 40.81% Conservative: 54
Best Local Similarity: 23.99% Mismatches: 100
Query Match: 4.77% Indels: 90
DB: Gaps: 18
US-09-807-459-2 (1-458) x CNS02TD6 (1-897)
Qy      192  ThrGIuLeuPheGIuTyrGIuIleLysArgAlaLeuLysGIu-----IleIleArg 208
          |||||
Db      29  ACTGCACATTTGGATGGGAGTAGATATTCCTCATGATGCAGGGCTATGAGTCTATTG 88
          |||||
Qy      209  SerAsnLeuPro-LeuAspIleGIuThrGIuIleSerValSer----- 222
          |||||
Db      89  TTCACCTCTCCACATCGACTGTGATCTGACGAGAGTAGTCCCTCGAGAAAGAACCC 148
          |||||
Qy      223  -----ArgLeuGIuIleIleThrSerSerTy 231
          |||||
Db      149  TGAATCTCGAAAAATTTGATTCGACACTTGTTCAGAAACTCGAGAAATTTGAAGCTTGCT 208
          |||||
Qy      231  rLysAspTyr---MetAspThrGIuIleProAla-----LeuProLy 244
          |||||
Db      209  GACGAATTATTCAGCAGACTTGACCTTCGACATCGAGAGATTACAGCATCTGCTGAA 268
          |||||
Qy      244  sPhe-----AlaLysArgPheSerLeuMetVal----- 254
          |||||
Db      269  ACTAGAGACGACGAAAGGAGAGTAGCGACGCGAGAGAGAGCCACCAAGTCAGATGCG 328
          |||||
Qy      255  -----GIuArgLeuLeuAlaThrValAlaGIuTyrValAspThrProTyrTyrLy 271
          |||||
Db      329  TCGAAGAAATACGACCTCTTCAGAGCATCGCC-----GACGTCCAAAACACAGA 379
          |||||
Qy      271  sLysTyrPyrIleLysLysLysAsnPheMetValAsnArgValPheIleProThrLysLy 291
          |||||
Db      380  GAAACAGCTGAAAAACCTCAACGCTGTGACAGAGAACATGATG----- 422
          |||||
Qy      291  sPhePheAsnLysGIuIleArgLysProSerLysAlaLeuLysGIuLysVal----- 308
          |||||
Db      423  -----GGAAACCTCACCGAGAGAGACGATCCCTCGAAGAAAAAGTAAACACAGA 472
          |||||
Qy      309  SerThrAspThrLysAspLeuPheGIuAsnLys-----IleGI 321
          |||||
Db      473  CGAAGACGACGCGAGAACCTATTCCTCGACGACAGATTKCAACATTTTAAATCA 532
          |||||
Qy      321  yGIuGIuLysThrValAspPhePheAsnLysGIuIleArgAspPro---SerLysAlaLeuLy 340
          |||||
Db      533  AAAAGACCGACGCGAGGAGAGAGAACAAATATACAGAGAGGAGGAGCTCACTCTA 592
          |||||
Qy      340  sGIuLysValSerAsnAspAlaLysAspLeuPheGIuAsnLysIleGIuGIuLysThr-- 359
          |||||

```





The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhae@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html). Plate: 287 row: H column: 12  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers

## FEATURES

## source

1. 794  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-287H12"  
/clone\_lib="RPCI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pTRABAC1; Site\_1: BamHI; Site\_2: BamHI. RPCI-24 mouse BAC library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 350 a 165 c 150 g 129 t  
ORIGIN

## Alignment Scores:

Pred. No.: 0.0319 Length: 794  
Score: 111.00 Matches: 62  
Percent Similarity: 39.00% Conservative: 39  
Best local Similarity: 23.94% Mismatches: 57  
Query Match: 4.71% Indels: 101  
DB: 12 Gaps: 14

US-09-807-459-2 (1-458) x BH098592 (1-794)

225 GlnHisIleThrSerSerTyrLysAspTyrMetLaspThr-GlnIleProAlaLeuProLy 244  
:::|||||::: ||| |||||||||  
Db 101 AAGATCTTACTATACAGAACCAACACGATCATCATCAAGAACGACATCTCCA-- 158  
244 sphenalalysarpheserleuMetValaGlnArgLeuLeuAlaThrValAlaGlyTy 264  
159 -----CTGCCCGACGACGACGAC 178  
Qy 264 rValAspThrProTrrPrrLysLysTrrPrrMetLysLeuLysAsnPhenMetValAsnAr 284  
:::|||||::: ||| |||||||  
Db 179 CCCGACACACCT-----GAAAGCTATACCTGATTTAAAGCATATCTCATGATGAT 232  
284 gValPheIleProThrLysLysPhePheAsnLysGluIleArgGluProSerLysAlaLe 304  
||| ||||| |||||  
Db 233 GGTA---GAGACATCAAGAGGACCTTAATTAACCTTAAGGAA-----AT 277  
Qy 304 uLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyTh 324  
:::|||||::: ||| |||||||  
Db 278 ACAGAGAAACACCTGCTTAACAGGTAGAAACATTAAGAG----- 317  
Qy 324 rValAspPheAsnLysGluIleArgAspProSerLysAlaLeuLysGluLysValaSe 344  
::: ||| |||||||  
Db 318 -----GAAGCAAAAATATCTTAATAAATGCAAGG 349  
Qy 344 rAsnAspAlaLysAspLeuPheGlu---AsnLysIleGlyGln 357  
||| ||| ||| ||||| |||  
Db 350 AAACACAAACAAACGATGATGAATTAATTAATCTTCAACACCTTAAGGAGGAGT 409  
357 ----- 357  
Db 410 AGACACAATAGAAACCAAGTGAAGGCAAGGCAAGCTGACATAGAAACCTTAGAAAGAA 469  
Qy 358 ---GlyThrValasp---PheIleAsnAsnGluIleArgAspProSerLysAlaLeuI 375

Db 470 ATCTGAACCATAGATGACGACATCAGCAACAGATACAGAGGTGAAGAGATC-- 527  
Qy 375 eArgLysValSerThrGlyValGluAspLeuPheGluAsn----- 388  
Db 528 -----TCAGGTGAGAGATGCGCATTAAGAAATCTCGGCACACAAATCAAGA 574  
Qy 389 -----LysIleGlyGlnGlyThrValAspPheIleAsnGluIleAr 403  
Db 575 AATGTGAATATGCAAAAGATCTATATCAAAATATTCAGGAAATGAGACACAAATGAG 634  
Qy 403 gAspProSerLysAlaLeuIleArgLysValTrrThrAlaAspAspLeuPheGluAs 423  
Db 635 AAGACCAAACTATGATGATATA-----GAAGTAGAT-----GAGAA 670  
Qy 423 nLysIleGlyGlnGlyThrValAspPheIleAsnLysGluIleArgAspProSer 441  
Db 671 TGAA-----GATTTCACCTCAAG-----GACCGGACA 698

RESULT 14 698 bp DNA linear GSS 16-FEB-2001  
A2766807  
LOCUS  
DEFINITION IM0564C11R Mouse 10kb plasmid UGSC1M library Mus musculus genomic  
clone UGSC1M0564C11 R, DNA sequence.  
ACCESSION A2766807  
VERSION A2766807.1 GI:12884256  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 698)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
JOURNAL CONTACT: Robert B. Weiss  
COMMENT University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunne@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
Plate: 0564 Row: C Column: 11  
Seq primer: CACACAGGAAACACCTATGACC  
Class: plasmid ends  
High quality sequence stop: 698.

## FEATURES

## source

1. 698  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UGSC1M0564C11"  
/clone\_lib="Mouse 10kb plasmid UGSC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative

of pMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 335 a 120 c 125 g 118 t

## ALIGNMENT SCORES:

Pred. No.:	0.0298	Length:	698
Score:	110.50	Matches:	50
Percent Similarity:	45.70%	Conservative:	35
Best Local Similarity:	26.88%	Mismatches:	56
Query Match:	4.68%	Indels:	45
DB:	12	Gaps:	11

US-09-807-459-2 (1-458) x AZ152840 (1-698)

```
OY 277 LeuylsAsnPhmetValAsnArgValPheileProThrLysLysPheAsn----- 294
      |||||.....:|||||
DB 6 TTTAAATCATATCTCTACATGATGAGACATTAAGAGACTTTAAATCACTCACTT 65
      |||||.....:|||||

OY 295 LysGluileArgGluProSer-----LysAlaLeuLysGluLysValSerThr 310
      |||||.....:|||||
DB 66 AAAGAAATACAGGAGACACACTGCTTAAGAGATTACAGTCTTAAGAAAGAAACAGCAAAAC 125
      |||||.....:|||||

OY 311 AspThrLysAspLeuPheGluAsnLysileGlyInGlyThrValAspPheAsnLys 330
      ::|||:::|||||
DB 126 ACATCCAAAGAGGTGATGAA-----ATGAACAA 155
      |||||.....:|||||

OY 331 GluileArgAspProSerLysAlaLeuLys-----GluLysValSerAsnAspAlaLys 348
      |||||.....:|||||
DB 156 ACCATATTATGACCAAAAGAGAGATGACACATTAAGAAAGAAAGAGAGCAAG 215
      |||||.....:|||||

OY 349 AspLeuPheGluAsnLysileGlyIn-----GlyThrValAsp---PheileAsn 365
      |||||.....:|||||
DB 216 CTGAGAGATGAAGACC---CTAGGAAGAAATCTGGAACCATGATGAGCATCGACAC 272
      |||||.....:|||||

OY 366 GluileArgAspProSerLysAlaLeuLysileArgLysValSerThrGluileGluAspLeu 385
      |||||.....:|||||
DB 273 AGATATCAAGAGATGAGAGAGATC-----TCAGGTACAGAGATGCC 317
      |||||.....:|||||

OY 386 PheGluAsn-----LysileGlyInGly 393
      |||||.....:|||||
DB 318 AAAGAGACATCTGCACAAAGCAAGCAAAATGCCAAATGCAATCTTAACCTAA 377
      |||||.....:|||||

OY 394 ThrValAspPheileAsnAsnGluileArgAspProSerLysAlaLeuileArgLysVal 413
      ::|||:::|||||
DB 378 AACATCCAGTATGTCAGAGACACATGAGAACCAAT-----CTACGGATTAATA 428
      |||||.....:|||||

OY 414 TyrThrGluAlaAspAspLeuPheGluAsnLysileGlyInGlyThrValAspPheile 433
      ::|||:::|||||
DB 429 GGAGTAGATGAGATGAGATGATTTCAACTTAA-----GGGCCAGCAACTTCTTC 479
      |||||.....:|||||

OY 434 AsnLysGluileArgAsp 439
      |||||.....:|||||
DB 480 AACAAATTTATGAAGAA 497
      |||||.....:|||||
```

RESULT 15 324 bp DNA linear GSS 28-AUG-2000  
AZ152840

LOCUS SP\_0034\_B2-G06-SP6 Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus

ACCESSION AZ152840  
VERSION AZ152840.1 GI:8304741

KEYWORDS GSS.  
Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus  
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
Echinoidea; Euechinoidea; Echinacea; Echinoida;

## REFERENCE

Strongylocentroidae: Strongylocentrotus.  
1 (bases 1 to 324)  
Cameron,R.A., Malaitas,G., Rast,J.P., Martinez,P., Blondi,T.R.,  
Swartzell,S., Wallace,J.C., Poultka,A.J., Livingston,B.T., Wray,  
G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and  
Hood,L.

## TITLE

A sea urchin genome project: Sequence scan, virtual map, and  
additional resources

## JOURNAL

Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

## MEDLINE

20402566

## COMMENT

Contact: Cameron, RA, Davidson, EH, Hood, L  
Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu  
Plate: 34 row: N column: 12  
Seg primer: SP6  
Class: BAC ends  
High quality sequence stop: 324.

## FEATURES

source

1..324

Location/Qualifiers

/organism="Strongylocentrotus purpuratus"  
/db\_xref="taxon:7668"

/clone="Plate=34 Col=12 Row=N"

/clone\_lib="Strongylocentrotus purpuratus, purple sea  
urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BACs3.6; BAC clones in E-Coli  
DH10B"

## BASE COUNT

81 a 95 c 91 g 57 t

## ORIGIN

## ALIGNMENT SCORES:

Pred. No.:	0.0135	Length:	324
Score:	109.00	Matches:	30
Percent Similarity:	45.13%	Conservative:	21
Best Local Similarity:	26.55%	Mismatches:	50
Query Match:	4.62%	Indels:	12
DB:	12	Gaps:	3

US-09-807-459-2 (1-458) x AZ152840 (1-324)

```
OY 294 AsnLysGluileArgGluProSerLysAlaLeuLysGluLysValSerThrAspThrLys 313
      |||||.....:|||||
DB 17 AACCAAGCCATGAGACTCCAGCTCCAGCCGCTGCTGTCGATGAGGAGACCCGA 76
      |||||.....:|||||

OY 314 AspLeuPheGluAsnLysileGlyInGlyThrValAspPheAsnLysGluileArg 333
      |||||.....:|||||
DB 77 GATTCGATGAGAACCCGATATCTTTGGGGAT-----AACCAAGCCATCAAG 124
      |||||.....:|||||

OY 334 AspProSerLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsn 353
      ::|||:::|||||
DB 125 ACTCCAGTCAACCTCGCCGCTGCTGCGCAGGGAGCCGGAATCGATGGAAC 184
      |||||.....:|||||

OY 354 LysileGlyInGlyThrValAspPheileAsnAsnGluileArgAspProSerLysAla 373
      ::|||:::|||||
DB 185 CGGTATCTTCGGGGAC-----AACCAAGCCATCAAAACCCCAAGTCCAGGC 232
      |||||.....:|||||

OY 374 LeuileArgLysValSerThrGluAlaGluAspLeuPheGluAsnLysileGlyInGly 393
      ::|||:::|||||
DB 233 TCGGGCGGTCTCTGCGATGGAGACCGCAGAGAGATGAGAACACGATATCTTCGGGG 292
      |||||.....:|||||

OY 394 ThrValAspPheileAsnAsnGluileArgAspProSer 406
      |||||.....:|||||
DB 293 GAT-----AACCAAGCCATCGAGACCCCAAGT 319
      |||||.....:|||||
```

## RESULT 16

AZ988471 642 bp DNA linear GSS 27-APR-2001

LOCUS 2M0271021F Mouse 10kb plasmid U06C2M library Mus musculus genomic

ACCESSION AZ988471  
clone U06C2M0271021 F, DNA sequence.

VERSION A2988471.1 GI:13859698  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 REFERENCE 1 (bases 1 to 642)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0271 row: 0 column: 21  
 Seq primer: CATTGTAAACGACGCCACT  
 Class: plasmid ends  
 High quality sequence stop: 642.  
 Location/Qualifiers  
 1. 642  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUC2M0271021"  
 /clone\_lib="Mouse 10kb plasmid UUC2M library"  
 /sex="Female"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42nv. Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 286 a 137 c 116 g 103 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.0391 length: 642  
 Score: 109.00 Matches: 61  
 Percent Similarity: 41.77% Conservative: 43  
 Best Local Similarity: 24.50% Mismatches: 77  
 Query Match: 4.62% Indels: 68  
 DB: 12 Gaps: 14  
 US-09-807-459-2 (1-458) x A2988471 (1-642)  
 QY 225 GlnHisIleTherSerSerrTyrLysAspTyrMetAspThrGlnIleProAlaLeuProLys 244  
 Db 19 AAGAAATCTACTTAACAAACCAAGCACCTACACATCATCAGAACCCAGACATCC-- 75  
 QY 245 PheAlaIysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValAlaGlyTyr 264

Db 76 -----ACTTGCCAGTCGAGGCAC 96  
 QY 265 ValAspThrProTrrTyrLysTrrTyrMetLysLeuLysAsnPhaMetValAsnArg 284  
 Db 97 CCCAAACACCC-----GAAACCTATACCCGGATTAAAGCCTATTCATC---ATG 147  
 QY 285 ValPheIleProThrLysLysPhePheAsn-----LysGluIleArgGluProSer 301  
 Db 148 ATGTTAAGAGCATCAAGAAAGGACATTAAATCACTTAAGAAATACAGAGACACACT 207  
 QY 302 -----LysAlaLeuLysGluLysValSerThrLysAspLeuPheGlu 317  
 Db 208 GCTAAATAGTAGAAGACATTAAAGAGAGAGACACAAAATCCCTTAAGAAATTGACAGAA 267  
 QY 318 AsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspProSerLys 337  
 Db 268 AACCAACCAACACG-----GTGATGGAATTGAATTAACCATTCACAAAC----- 312  
 QY 338 AlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGlu----- 352  
 Db 313 --CTAAAGGAGAGTAGACACAAATTAAGAAACCCAAAGTGAGCAACGCTGAGATA 369  
 QY 353 -----AsnLysIleGlyGln---GlyThrValAsp---PheIleAsnArgGluIle 367  
 Db 370 GAAACCTTAGAACACATAGCAAAATCTGGAACCTACATGCGACATCAGCAACAGATA 429  
 QY 368 ArgAspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGlu 387  
 Db 430 CAACAGATGCAAGAGACAGATC-----TCAGGTGAGAAAGANTTCATGAGAA 474  
 QY 388 AsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 407  
 Db 475 AAC-----ATCGGACCAACCAATCAAGAAATACAAAA 507  
 QY 408 AlaLeuIleArgLysValTrrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGln 427  
 Db 508 TCC-----AAAAGATCTCAACTCAA----- 528  
 QY 428 GlyThrValAspPheIleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLys 447  
 Db 529 ---AACATCCAGGAATTCAGACACATGACAGACCAATGCAACCAATGATATACGAGTA 585  
 QY 448 ValSerThrGluAlaAspAsnLeuLeu 456  
 Db 586 GATGAATGAAATGAAATTTTCGACTCATATA 612  
 RESULT 17  
 BB230024 700 bp mRNA linear EST 23-OCT-2001  
 LOCUS BB230024 RIKEN full-length enriched, 3 days neonate thymus Mus  
 DEFINITION musculus cDNA clone A630024K10 3', mRNA sequence.  
 ACCESSION BB230024  
 VERSION BB230024.2 GI:16354349  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota: Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 700)  
 AUTHORS Arakawa, T., Carlini, P., Fukuda, S., Furuno, M., Hanagaki, T., Harra, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
 JOURNAL Unpublished (2001)  
 COMMENT On Jul 3, 2000 this sequence version replaced gi:8900669.  
 Contact: Yoshinori Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute







Alignment Scores:

Score:	108.00	Length:	996
Percent Similarity:	36.56%	Matches:	62
Best Local Similarity:	22.22%	Conservative:	40
Query Match:	4.58%	Mismatches:	107
		Indels:	70
		Gaps:	11

US-09-807-459-2 (1-458) x CNS06HIG (1-996)

```

QY 220 SerValSerAlaGluGlnHisIlePheSerSerTyrLys-----AspTyr 234
Db 975 AGTGTATTGCACTTCTCAAAAAGCAAGCGATGATGCGCTTCATACACTTACACT 916
QY 225 MetAspThrGlnIleProAlaLeu----- 242
Db 915 ATTGAACGGAACATCCACATTAATCTTACCTGATTCACAGACTATTTTATCATTA 856
QY 243 -----ProLysPheAlaLysArgPheSerLeuMetValValGlnArg 256
Db 855 AATGCCATTGATTAGCAACAGAAATATGAAAAAGCAATGATATGTGCTGATTCT 796
QY 257 LeuLeuAlaThrValAlaGlyTyrVal---AspThrProTyrTyrLysTyrPyrMet 275
Db 795 ATCATCTGTAAACACTTAAACATTGCCAAGATTGAAATGAGAACAGCCCTTAGACT 736
QY 276 LysLeu-----LysAsnPheMetValAsnArgVal 285
Db 735 AAGTTAGTTACGATCGAAGGGCTTTATTCATTAAGCCGGTTAATGACAGGTGTATA 676
QY 286 PheIleProThrLysLysPhePheAsnLysGluIleArgGluProSerLysAlaLeuLys 305
Db 675 TCGGGGGAACCTAATATATAGATGGACAAAGAGATATCAAAAGCTTAATGCTTGAAG 616
QY 306 GluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGly----- 321
Db 615 GACAAGTTCGATATCAAAATCGACAGCTTTCACACAGCAAGCTTAGTCCATTAGG 556
QY 322 -----GlnGlyThrValAspPheAsnLysGluIleArgAspProSerLys 337
Db 555 GCCAGGAGGATCGAAMACAGTGTTCCTATGATTCGATATA-----GCA 508
QY 338 AlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGly--- 356
Db 507 AGTTTGAACTCAGATGACTCAACAAAGCGTTCCTTGATGAGATTAATATAGATA 448
QY 357 GlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArg 376
Db 447 AATATACCATATGACCTGATGATATGATTCACCAACAAAGATACCGAATTGGAGAG 388
QY 377 LysValSerThrGlnLysAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp 396
Db 387 AACTTA-----GAACATTTGGAAAAAGCTAAG----- 361
QY 397 PheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValTyrThrGlu 416
Db 360 -----AATGAAATTGAGAACAGAGAACAGCGTTTACAAAGGAGATTTCACAAAGA 310
QY 417 AlaAspAspLeuPheGluAsnLysIleGlnGlyThrValAspPheIleAsnLysGlu 436
Db 309 -----TTTACAGAGAAATCGG-----TTCTCATTTAAAGAG 277
QY 437 IleArgAspProSerLysAlaLeuIleArgLysValSerThrGlnAlaAspAsnLeu 455
Db 276 TATGAAGACTCATTCGCGTGAATTAATGAGACAAATCAATGCAACAAATTG 220

```

RESULT 21  
CNS07C18 1154 bp DNA linear GSS 08-JUL-2001  
LOCUS T3 end of clone XBC0AA001D03 of library XBC0AA from strain CBS 767  
DEFINITION of Debaryomyces hanseni1, genomic survey sequence.  
ACCESSION AL438594  
VERSION AL438594.1 GI:12222007

KEYWORDS GSS.  
SOURCE Debaryomyces hanseni1.  
ORGANISM Debaryomyces hanseni1.  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
REFERENCE  
1 (bases 1 to 1154)  
Lepingle, A., Casaregola, S., Neuvéglise, C., Bon, E., Nguyen, H.,  
Artiguenave, F., Winkler, P. and Galliard, C.  
Genomic exploration of the hemiascomycetous yeasts: 14.  
Debaryomyces hanseni1 var. hanseni1  
FEMS Lett. 487 (1), 82-86 (2000)  
20584724.  
2 (bases 1 to 1154)  
Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,  
Boulet, F., Kuhn, M., Bon, E., Brotier, P., Casaregola, S.,  
de Montigny, J., Dujon, B., Durand, P., Lepingle, A., Llorente, B.,  
Malpertuy, A., Neuvéglise, C., Ozier, K., Kalogeropoulos, O., Potier, S.,  
Saurin, W., Tekala, F., Toffano-Nicohe, C., Wesolowski, L., Winkler, P. and Weissenbach, J.  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
FEMS Lett. 487 (1), 3-12 (2000)  
20584711.  
3 (bases 1 to 1154)  
Genoscope.  
Direct Submission  
Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage,  
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :  
sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hanseni1 var. hanseni1, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
the other extremity of this insert.  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.  
location/Qualifiers  
1..1154  
/organism="Debaryomyces hanseni1"  
/strain="CBS 767"  
/variety="hanseni1"  
/db\_xref="taxon:4959"  
/clone="XBC0AA001D03"  
/clone\_11b="XBC0AA"  
/note="end : T3"  
misc\_feature  
complement(<6..>1115)  
/note="similar to Saccharomyces cerevisiae ORF YD074C [ weak similarity to spindle pole body protein NUP1 ]"  
/evidence="not\_experimental"  
BASE COUNT 268 a 226 c 149 g 507 t 4 others  
ORIGIN

Alignment Scores:

Score:	108.00	Length:	1154
Percent Similarity:	40.19%	Matches:	65
Best Local Similarity:	20.57%	Conservative:	115
Query Match:	4.58%	Mismatches:	74
		Indels:	14
		Gaps:	14

US-09-807-459-2 (1-458) x CNS07C18 (1-1154)

```

QY 172 LysThrTyrThrAsnValAspGluPheGlyAlaSerPheAsnLysLeuSerPheThr 191
Db 953 AAGACTCTTCACAGAAATTGACGA-----AGTTAAC 921
QY 192 ThrGlyLeuPheGlyTyrPglyTlleLysArgAlaLeuLysGlnIleIleArgSerAsnLeu 211
Db 920 AACGATCG-----GAGTGAAGAGAGAGCTTCGAGACAGTAAATCTGAGTA 870

```



```

Oy 212 Prolenasp-----TleclYthrgluHnlserval-----SerargleuGlnHnlsile 227
    |||
    |||
    |||
Db 869 AATAAGCACACAAACGGTGAATAAATGAGATTCGTACAGACATGAACAGATTGAAGGATTA 810
Oy 228 ThrsrSertYlYsAsPtyrMetAspThrGlnlleProAlaleuProLysPheAlaLys 247
    |||
    |||
    |||
Db 809 AGTTGAGAAATGAAAGAG-----792
Oy 248 ArgPheSerMetValValGlnArgleuLeuAlaThrValAlaGlyTyrValAspThr 267
    |||
    |||
    |||
Db 791 -----TTGAAAAGCCATTAATAGTTATTAAGAAAGCACTGATCATCATCAATAC 741
Oy 268 ProTpyrTyrLys-----LysTpyrMetLysLeu-----277
    |||
    |||
    |||
Db 740 AACTATGATTAAGTAGTAAGAGAGATTTACAGTTATCTGATTAACACTAGAACTTGAT 681
Oy 278 LysAsnPhmeMetValAsnArgValPheIleProThrLysLysPheAsnLysGluile 297
    |||
    |||
    |||
Db 680 GAAAACGACTGTGATTAATCCGTTTCGTATCAAGAAATTCGACACATTAACAACAAATG 621
Oy 298 ArgGluProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGlu 317
    |||
    |||
    |||
Db 620 AACGAAAGCGTAATRAATTTTCAAAAAATTATGACACACACTGTAATTAAGTTAAATGAG 561
Oy 318 AsnLysIleGlyGlnGlyThrVal---AspPheAsnLysGluileArgAspProSer 336
    |||
    |||
    |||
Db 560 TTAGAAATTAAGCAACCAATCTGAAGCCTTACTTACACAGAAATTAAGGAGGAGAAAT 501
Oy 337 LysAlaLeuLysGluLysValSerAsnAlaLysAspLeuPheGluAsnLysIleGly 356
    |||
    |||
    |||
Db 500 GAAACTCTAAACCAACAAATTACAGAAAGCGAAACGATTTGTC-----456
Oy 357 GlnGlyThrValAspPheLeuAsnGlnlleArgAspProSerLysAlaLeuileArg 376
    |||
    |||
    |||
Db 455 -----AGAAATTCGAACCTCCAGACAGATGAGTTACTATCT 423
Oy 377 LysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp 396
    |||
    |||
    |||
Db 422 AAACAAACAACTTGAAGAGCTGAT---TTTGAAATCAACACTACTATCATCATTAAT 366
Oy 397 PheileAsnAsnGluileArgAspProSerLysAlaLeuileArgLysValTyrThrGlu 416
    |||
    |||
    |||
Db 365 AAATGAAATAGGCTCTGATGAAGAATTAATGCTTGAAAGCGAAAGCATGAAACT 306
Oy 417 AlaAsp-----AspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp 431
    |||
    |||
    |||
Db 305 GGTGATTAATAGTAGATTAAGCGAATTAATTCGAAGATGATGATTCAA---CGATTTAAC 249
Oy 432 PheileAsnLysGluileArgAspProSerLysAla-----443
    |||
    |||
    |||
Db 248 ATGTTAAATAGCGAAATCAAGAAATAGAGCAAGCATTTTCAAGAAACGAGAGAAACTCG 189
Oy 444 LeuileArgLysValSer---ThrGluAlaAspAsnLeuLeuGluLys 458
    |||
    |||
    |||
Db 188 CTATCTAAGTAAATGAGTGTACTACACCAAGAAATATGCTTAAAAAG 141

```

## RESULT 22

```

A2431579 642 bp DNA linear GSS 03-OCT-2000
LOCUS 1M0216006R Mouse 10kb plasmid UUCG1M 1library Mus musculus genomic
DEFINITION clone UUCG1M0216006 R, DNA sequence.
ACCESSION A2431579
VERSION A2431579.1 GI:10555592
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.

```

REFERENCE 1 (bases 1 to 642)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 JOURNAL plasmid inserts  
 COMMENT Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 1000 Std Error: 0.00  
 Plate: 0216 row: 0 column: 06  
 Seq primer: CACACGGAACACAGCATACAC  
 Class: plasmid ends  
 High quality sequence stop: 642.  
 Location/Qualifiers  
 1..642  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUCG1M0216006"  
 /clone\_1lb="Mouse 10kb plasmid UUCG1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="vector: PMD42nv. Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pMD42 (g14732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 adapted vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

## FEATURES

source

BASE COUNT 294 a 125 c 116 g 107 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 0.0585 Length: 642  
 Score: 107.50 Matches: 58  
 Percent Similarity: 44.78% Conservative: 32  
 Best Local Similarity: 28.86% Mismatches: 66  
 Query Match: 4.56% Indels: 45  
 DB: 12 Gaps: 12

US-09-807-459-2 (1-458) x A2431579 (1-642)

```

Oy 263 glyTyrValAspThrProTpyrTyrMetLysLeuLysAsnPhmeVal 282
    |||
    |||
    |||
Db 96 GGGACACCCCAACACACC-----GAAAAGATAGACCCGATTTAAAGCATATCTCAG 149
Oy 283 AsnArgValPheIleProThrLysLysPheAsn-----LysGluileArgGluPro 300
    |||
    |||
    |||
Db 150 ATGATGTAGAGAGACATCAAGAACTTAACTCACTTAAGAAATACAGAGAGAC 209
Oy 301 Ser-----LysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPhe 316
    |||
    |||
    |||
Db 210 ACTGCTAAACAGATAGACATCTTAAGAGAAAGCAAGAAATCCCTTAAAGATTTGAG 269
Oy 317 GluAsnLysIleGlyGlnGlyThrValAspPheAsnLysGluileArgAspProSer 336
    |||
    |||
    |||
Db 270 GAAACATGACCAACACAG-----GTGATGGAATTAATTAACCAATCAACAGC-----317

```



ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 478) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert length: 10000 Std Error: 0.00 Plate: 0271 row: A column: 02 Seq primer: CACACAGCAACACGCTATGACC Class: plasmid ends High quality sequence stop: 478. Location/Qualifiers 1..478 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="UUGC1M0271A02" /clone_lib="Mouse 10kb plasmid UUGC1M library" /sex="Male" /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-" /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114[gblAF129072.1]), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT	100 a 95 c 92 g 191 t
ORIGIN	
Alignment Scores:	
Pred. No.:	0.0721 Length: 478
Score:	105.00 Matches: 46
Percent Similarity:	44.38% Conservative: 33
Best Local Similarity:	25.84% Mismatches: 55
Query Match:	4.45% Indels: 44
DB:	12 Gaps: 8
US-09-807-459-2 (1-458) x AZ462925 (1-478)	
OY	197 TRPGlyIleLysArgAlaLeuLysGlnIleIleArgSerAsnLeuProLeuAspIlegly 216 :::      111 111      :: 111         111         :::
Db	441 TTGTGGTGTCCAAAGAGTGCTGCACACAGGTTTCACA-----CTTGTAAGAGT 394
OY	217 ThrGluHisSerValSerArgLeuGlnHisIleThrSerSerTyrllysAsptyrMetasp 236 111 111  111:: 111 111:: 111         :::         :::
Db	393 ACAAGCTCCACGCCACAGTATTATCATCTAACACACAGAGATGACCAAAATGGTGAAG 334

FEATURES	SOURCE
Db 237	ThrgInIIProAlaLeuProLysSphenalLals-----ArgPheSerLeu 251
Db 333	GCAAAAGGTAAGATCTTACCAACAGAAACCAAGACTACATGGCATCATGACCACTA 274
Db 252	MetValValGlnArgLeuLeuAlaThrValAlaGlyTyrValAspThrProTyrLys 271
Db 273	CTTTACCAACAGCGAGTCTCTG-----GATACCCCAACATACCGG 235
Db 272	LysTyrTyrMetLys--LeuLysAsnPhenMetValAsnArgValPheIleProThrLysL 291
Db 234	AAAACAGCATTCAGATTAAATATCATATCATATGATGCTTGTA----- 191
Db 291	YSPhePheAsnLysLeuLeuArgLysProSerLysAlaLeuLysLeuLysValSerThrA 311
Db 190	-----GAGAAATTAAAGAAACATTAATTAATT 163
Db 311	sPTThLysAspLeuPheGlnAsnLysLysLysGlyGlnGlyTyrThrValAspPheAsnLysG 331
Db 162	GCCTTAAAGAAATACAGAGAAACACAGTTAAACAGCTGAAGTC-----CCTAAAG 112
Db 331	IuIleArgAspProSerLysAlaLeuLysGlyLysValSerAsnAspAlaLysAspLeuP 351
Db 111	AG-----GAAGCAACAAAATCCCTTAAAGATTACAGAAACAAACACAAACAGCTGA 58
Db 351	heglu-----AsnLysIleGlyGlnGlyTyrThrValAspPheIleAsnAsnLys 366
Db 57	AGGAATTGACAAA-----ACCATCCACTACTCTGAAAGAGAG 20
RESULT 25	
LOCUS	BM273998 532 bp mRNA linear EST 20-DEC-2001
DEFINITION	PfESToa62d10.y1 Plasmodium falciparum 3D7 gametocyte cDNA library
ACCESSION	Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.
VERSION	BM273998
KEYWORDS	BM273998.1 GI:17967300
SOURCE	EST.
ORGANISM	Plasmodium falciparum 3D7.
REFERENCE	Plasmodium falciparum 3D7
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
	1 (bases 1 to 532)
	Tang R., Cole R., Chakrabarti D., Haywood R., Clifton S., Pape D.,
	Marrar M., Hillier L., Martin J., Wylie T., Dante M., Theising B.,
	Bowers Y., Gibbons M., Ritzer E., Bennett J., Jones E., Konko I.,
	Tsagaris I., R., Richey J., Wadkins J., Kennedy S., Levinso D.,
	Waterston R., Wilson R., and Sibley D.
TITLE	WashU Plasmodium EST Project
JOURNAL	Unpublished (2001)
COMMENT	Contact: L. David Sibley
	WashU Plasmodium EST Project
	Washington University School of Medicine
	444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
	Tel: 314 286 1800
	Fax: 314 286 1810
	Email: est@watson.wustl.edu
	Library was constructed by R. Haywood. DNA sequencing by:
	Washington University Genome Sequencing Center for information on
	obtaining a clone please contact: L. David Sibley
	(sibley@borcim.wustl.edu), Washington University
	Seq primer: -40up from Glbco
	High quality sequence stop: 422.
	Location/Qualifiers
	1..532
	/organism="Plasmodium falciparum 3D7"
	/db_xref="taxon:36329"
	/clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA
	library"
	/dev_stage="gametocyte (stage III-V)"
	/lab_host="DH10B (Genesig, Invitrogen, Inc.)"
	/note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2:
	XhoI; The library was constructed by R Haywood. cDNAs were
	synthesized from gametocyte poly(A)+ RNA by oligo d(T)
	priming, size-selected and directionally cloned into the



```

QY 302 Lysala-----LeuLysGluLysValSerThrAspThrLysAspLeu 315
DB 318 AACACTGCTAAACAGGTAGACACATTTAAAGAGAAACCAAAATCCCTTAAAGATTG 259
QY 316 PheGluAsnLysIleGlyInglYthrValAspPheAsnLysGluIleArgAspPro 335
DB 258 GAGGAAACACACAAACCAACAG-----GTGATGGAATTCATTAATTCATCAAGAC--- 208
QY 336 SerLysAlaLeuLysGluLysValSerAsnAspAlaLys----- 348
DB 207 -----CTAAAGAGGAGTAGACACAAATTAAGAAACCAAGAGGCAACGCTG 157
QY 349 -----AspLeuPheGluAsnLysIleGlyInglYthrValAsp---PheIleAsnAsn 365
DB 156 GAGATAGAAACCTAGACAAAGAAA-----TCTGGAACCATAGATCGAGCATCAGCAAC 103
QY 366 GluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeu 385
DB 102 AGAATTCAGAGATGAGAGAGATC-----TCAGGTGCGAGAAAGATTCC 58
QY 386 PheGluAsnLysIleGlyInglYthrValAspPheIleAsnAsnGluIleArgAspPro 405
DB 57 ATAGAGAAC-----ATCGGCAACACAAATCAAGAAAG 25
QY 406 SerLys 407
DB 24 GAAAG 19

RESULT 27
LOCUS BM168451 644 bp mRNA linear EST 04-DEC-2001
DEFINITION ESF570974 PYBS Plasmodium yoelii yoelii cDNA clone pYCPA68 5' end,
RNA sequence.
ACCESSION BM168451
VERSION BM168451.1 GI:17301683
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
REFERENCE 1 (bases 1 to 644)
AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,
Fraser,C.M. and Carucci,D.J.
Plasmodium yoelii EST project at TIGR
Unpublished (2001)
TITLE JOURNAL
COMMENT Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.
FEATURES
Source
1..644
Location/Qualifiers
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCPA68"
/clone.lib="PYBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cbyJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of

```

```

5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

BASE COUNT 310 a 69 c 109 g 156 t
ORIGIN

Alignment Scores:
Pred. No.: 0.15 Length: 644
Score: 104.00 Matches: 48
Percent Similarity: 46.328 Conservative: 40
Best Local Similarity: 25.268 Mismatches: 66
Query Match: 4.41% Indels: 36
DB: Gaps: 11

US-09-807-459-2 (1-458) x BM168451 (1-644)
QY 276 LysLeuLysAsnDpHekeValAsnArg-----ValPheIlePro 288
DB 37 AAATTAATAAATAGAGTAGAATAAACCAGCAATTAATAAATGACATGCAATTA 96
QY 289 ThrLysLysPhePheAsnLysGluIleArgGluProSerLysAlaLeuLysGluLysVal 308
DB 97 ACCGAAATAAATAAATAAT--GAAGTATCGAATTAACCGAAATAAATAAATAATGAAGTA 153
QY 309 SerThrAspThrLysAspLeuPheGluAsnLysIleGlyInglYthrValAspPhe 328
DB 154 TCGAATAAATAACGAAATAAATA--AAATGAGATTCGATTAACCGAAATAAATAA 210
QY 329 AsnLysGluIleArgAspProSerLysAlaLeuLysGluLysValSerThrGlyAla 348
DB 211 AAT--GAAGTACCTAATAAACCAGAAACAAATTAACACACAGTA--GATGATATTA 264
QY 349 AspLeuPheGlu-----AsnLysIleGlyInglYthrValAspPheIleAsnAsn 365
DB 265 GATTAATATTAAAGATAGTGGAACCTCTTGATGAGATCAATCCCAATCGAATA 324
QY 366 GluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGlu----- 383
DB 325 ACTATTAAAGATACCTGTGTGATACATAGGGGATATGCTCTGATTTGAAAAACA 384
QY 384 ---AspLeuPheGluAsnLysIleGlyInglYthrValAspPheIleAsnAsnGluIle 402
DB 385 TCTGATGTAGTTAAAGAAAAAGTA-----ATTGAACATATTATTATTC--- 429
QY 403 ArgAspProSerLysAlaLeuIleArgLysValIleThrGluAlaAspLeuPheGlu 422
DB 430 --GATATTGTCCTCCGCCCTTAAGAGAGATTGT----- 462
QY 423 AsnLysIleGlyInglYthrValAspPheIleAsnLysGluIleArgAsp----- 439
DB 463 AATATGTCGTGTAATAATATTACTCAATTAATGAAAAAGATTAAGATGAAGATCC 522
QY 440 -----ProSerLysAlaLeuIleArgLys 447
DB 523 GATTAAGTCAAAAGCATTTTAAAAA 552

RESULT 28
LOCUS BH126086 715 bp DNA linear GSS 19-JUL-2001
DEFINITION RPCI-24-289K13.TJ RPCI-24 Mus musculus genomic clone RPCI-24-289K13
, DNA sequence.
ACCESSION BH126086
VERSION BH126086.1 GI:14969598
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 715)	
AUTHORS	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Aktenre,B., Levin,M., Tsegaye,G., Geer,K., Krol,W., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE	Mouse BAC End Sequences from Library RPCI-24
JOURNAL	Unpublished (1999)
COMMENT	Other-GSS: RPCI-24-289K13.TV Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@tigr.org). Clones may be purchased from BACPAC Resources ( <a href="http://www.chori.org/bacpac/orderingframe.html">http://www.chori.org/bacpac/orderingframe.html</a> ). BAC end plate: <a href="http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html">http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html</a> Page: 289 row: K column: 13 Seq primer: SP6 Class: BAC ends.
FEATURES	
source	Location/Qualifiers 1..715 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="RPCI-24-289K13" /clone_11b="RPCI-24" /sex="Male" /cell_type="Spleen/Brain" /note="Vector: pTRABAC1; Site.1: BamHI; Site.2: BamHI; RPI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."
BASE COUNT	126 a 152 c 159 g 278 t
ORIGIN	
Alignment Scores:	
Pred. No.:	0.202 Length: 715
Score:	103.50 Matches: 46
Percent Similarity:	45.86% Conservative: 37
Best Local Similarity:	25.41% Mismatches: 61
Query Match:	4.39% Indels: 37
DB:	12 Gaps: 10
US-09-807-459-2 (1-458) x BH126086 (1-715)	
Oy	216 G1yThrgLuhIsSerValSerArgrLeuGlnHisIleThSersertyrLysAspTyrMet 235
Db	468 GGTTACACAGATGGGGAAGGCAACGTAACAACTTGTCTAACAGAAACCAACACACTCA 409
Oy	236 AsPthrgInIleProAlaLeuProLysPheAlaLysArgPheSerLeuMetValValGln 255
Db	408 CCATCATCAACAACCCGGGCACTCCCACTCAAGCCAGT----- 373
Oy	256 ArgLeuLeuAlaThrValAlaGlyTyrValAspPthProTPrtyrLysLysTPrtyrMet 275
Db	372 -----CCTGGATTAACCCCAACACACACC-----GAAAGACCAACACTTG 337
Oy	276 LysLeuLysAsnProMetValAsnArgValPheIleProThrLysLysPheAsn--- 294
Db	336 GATTAAATATATATTCATCATGATGCTGGTAGAGACATCAAGAGACTTTAATACTCA 277
Oy	295 ---LysGluIleArgGluProSer-----LysAlaLeuLysGluLysValSer 309
Db	276 CTTTAAAGAAATACAGAGAAACACTGCTAAAGAGGTGAAAGTCCCTTAAAGAAAAACAGAA 217
Oy	310 ThrAspThrLysAspLeuPheGlu---AsnLysIleGlyLysGlnGlyThrValAspPhePhe 328
Db	216 AACACACATTAACAGGATGATGGAAATTTGAACAAA-----ACCAACACAGACCTTA 169

QY	329	AsnLysGluIleArgAspProSerLysAlaLeuLysGluValSerAsnAspAlaLys	348
Db	168	AAAAAGGAGAA---GAAACAATAAAGAAATCCAAAGTAGACACACTTGGAGCTAGAA	112
QY	349	AspLeuPheGluAnLysIleGlyGlnGlyThrValAsp---PheIleAsnAsnGluIle	367
Db	111	ACCTGAGAAAGAMT-----TCTGGAAACCATAGATGGCAGCATCAGCAAGATA	61
QY	368	ArgAspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGlu	387
Db	60	CAGAAGATGAGAAAGATTAATC-----TCAGGTGCGAAGATTCCATAGAG	16
QY	388	Asn	388
Db	15	AAC	13
RESULT	29		
LOCUS	BG523384		
DEFINITION	BG523384	717 bp	mRNA
ACCESSION	BG523384		
VERSION	BG523384		
KEYWORDS	EST		
SOURCE	Stevia rebaudiana.		
ORGANISM	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; easterids II; Asterales; Asteraceae; Asteroideae; Heliantheae; Stevia.		
REFERENCE	1 (bases 1 to 717)		
AUTHORS	Brandt, J.E., Richman, A., Swanson, A.K. and Chapman, B.P.		
TITLE	Leaf ESTs from Stevia rebaudiana: A resource for gene discovery in diterpene synthesis		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Jim Brandt Genomics and Biotechnology Agriculture and Agri-Food Canada - SCFPFC 1391 Sandford St., London, Ontario, CANADA, N5V 4T3 Tel: 519 457 1470 Fax: 519 457 3997 Email: brandtj@em.agr.ca Seq primer: T3 promoter primer.		
FEATURES	Location/Qualifiers		
SOURCE	1..717		
	/organism="Stevia rebaudiana"		
	/strain="751/1501"		
	/cultivar="Landscape"		
	/db_xref="taxon:55670"		
	/clone_lib="Stevia field grown leaf cDNA"		
	/tissue_type="leaf"		
	/dev_stage="field grown, mid-size"		
	/lab_host="E. coli strain XL04"		
	/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from polyA+ enriched mRNA from field grown leaves. Mid-size actively growing leaves were collected and pooled from several plants and frozen immediately after harvesting in liquid nitrogen. The cDNA was prepared using an XhoI-poly(dT) linker-primer. An EcoRI adapter was ligated to the blunt end cDNA and the products were digested with EcoRI and XhoI enabling directional cloning into the lambda ZAP Express vector. The library was amplified using the host strain XL1-Blue MRF. Mass excision of the library was performed to obtain pBK-CMV phagemid clones in the host strain XL04. Single pass DNA sequencing was performed using the T3 promoter primer: 5' ATTAACTCTACTTAAAGGA 3'. This library was constructed by Alex Richman."		
BASE COUNT	261 a	90 c	161 g
ORIGIN			205 t

Pred. No.: 0.231 Length: 717  
 Score: 103.00 Matches: 46  
 Percent Similarity: 45.458 Conservative: 49  
 Best Local Similarity: 22.018 Mismatches: 74  
 Query Match: 4.37% Indels: 40  
 DB: 10 Gaps: 10

US-09-807-459-2 (1-458) x BG523384 (1-717)

QY 277 LeuLysAsnPhMetValAsnArgValPhe-----IleProThr 289  
 |||||  
 DB 14 TTAAGGAATTAAGTCAAGCAAGCTTAATCTTGGTGAATAATTAACAAAAGACCTTCCTTG 73  
 290 LysLysPhePheAsnLysGluLLeuArgGlu-----ProSerLysAlaLeuLysGluLys 307  
 |||||  
 DB 74 AAGGCTAAAGTCAATGAG 133  
 308 ValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhe 327  
 |||||  
 DB 134 ATTGAATAATTGAAAG 187  
 328 PheAsnLysGluLLeuArgAspProSerLysAlaLeuLysGluLysValSerAsnAspAla 347  
 |||||  
 DB 188 TTTGATCAAG 247  
 348 LysAspLeuPheGlu-----AsnLysIleGlyGlnGlyThr----- 359  
 |||||  
 DB 248 GTTAAAGCCCTTGAATCATCGGGTTCATGTCAGTCTCAAAATTAAGTCAAGTTCATG 307  
 360 -----ValAspPheIleAsnAsnGluLLeuArgAspProSerLysAla 373  
 |||||  
 DB 308 TCGGTAATTAAGCGTAAGTGTGATGAGAACTTAATGCAAGAAATTAATGAG----- 355  
 374 LeuLLeuArgLysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGly 393  
 |||||  
 DB 356 ATAATTCAG 412  
 394 ThrValAspPheIleAsnAsnGluLLeuArgAspProSerLysAlaLeuLLeuArgLysVal 413  
 |||||  
 DB 413 AAACCTTGAGCTAGTAAAGACACAAA---AATACCCCTAGTAGAGAAATCTAAGAAAGATG 469  
 414 TyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIle 433  
 |||||  
 DB 470 -----GAGCCTTACGACGACGATTAAGAGGCTTCTGCTGATGTCAG 514  
 434 Asn-----LysGluLLeuArgAspProSerLysAlaLeuLLeuArgLysValSer 449  
 |||||  
 DB 515 AGTTTCGACGAGCTTAAGAGAAATATGAAAGATTAACAGCTGACAGACTTCAGAACTTCG 574  
 450 ThrGluAlaAspAsnLeuGluLys 458  
 |||||  
 DB 575 GGTGATCTGATGAGAACTTGTATCAAA 601

RESULT 30  
 BH047254 791 bp DNA linear GSS 17-JUL-2001  
 LOCUS BH047254/c  
 DEFINITION RPI-24-242M13.TJ RPI-24 Mus musculus genomic clone RPI-24-242M13  
 , DNA sequence.  
 ACCESSION BH047254  
 VERSION BH047254.1 GI:14833603  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 791)  
 Zuo,S., Nierman,M., Malek,J., Shaltsman,S., Akintret,B., Levins,M.,  
 Tsugey,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorjgis,E.,  
 Russell,D., de Jong,P. and Fraser,C.M.  
 Mouse BAC End Sequences from Library RPI-24  
 Unpublished (1999)  
 Other GSSs: RPI-24-242M13.TV  
 Contact: Shaying Zhao

Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@ligr.org  
 Clones are derived from the mouse BAC library RPI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end  
 page: [http://www.ligr.org/tdb/bac-ends/mouse/bac\\_end\\_intro.html](http://www.ligr.org/tdb/bac-ends/mouse/bac_end_intro.html)  
 Plate: 242 row: M column: 13  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source location/Qualifiers  
 1..791  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPI-24-242M13"  
 /clone\_1db="RPI-24"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /note="Vector: pTRABAC1; site\_1: BamHI; site\_2: BamHI;  
 RPI-24 Mouse BAC library produced by Pieter de Jong. The  
 library was cloned in the pTRABAC1 cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."

BASE COUNT 136 a 144 c 163 g 348 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.27 Length: 791  
 Score: 103.00 Matches: 71  
 Percent Similarity: 41.16% Conservative: 43  
 Best Local Similarity: 25.63% Mismatches: 88  
 Query Match: 4.37% Indels: 75  
 DB: 12 Gaps: 17

US-09-807-459-2 (1-458) x BH047254 (1-791)

QY 199 IleLysArgAlaLeuLysGluLLeuArgSerAsnLeuProLeuAspIleGlyThrGlu 218  
 |||||  
 DB 708 ATTACGAGAAAGATTAAGGCAAAAATAAG----- 679  
 219 HisSerValSerArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGln 238  
 |||||  
 DB 678 -----ATCCTGCTAACAGAAATCAAGACCACTCACCATGAT---CAGAACACAGCA 631  
 239 IleProAlaLeuProLysPheAlaLysArgPheSerLeuMetValGlnArgLeuLeu 258  
 |||||  
 DB 630 CTCCTCAACCCCACTAGT----- 613  
 259 AlaThrValAlaGlyTyrValAspThrProTyrLysTyrTrpMetLysLeuLys 278  
 |||||  
 DB 612 -----CCTGGGATATCCCAACACACC-----GAAAGCTAGACTCAGATTAAAA 568  
 279 AsnPhMetValAsnArgValPheIleProThrLysLysPheAsn-----LysGlu 296  
 |||||  
 DB 567 GCATATCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508  
 297 IleArgGluProSer-----LysAlaLeuLysGluLysValSerThrAspThr 312  
 |||||  
 DB 507 ATACAGAGAGACCTCTTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 448  
 313 LysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheAsnLysGluLLeu 332  
 |||||  
 DB 447 AAAGATTCGAGAGAAAGCATGACCAACAG-----GTGATGGAATTTGAATTAACCATC 394  
 333 ArgAspProSerLysAlaLeuLys-----GluLysValSerAsnAspAlaLysAspLeu 350  
 |||||  
 DB 393 CAAGACATATAAAGGAGATGACACACAAATTAAGAAATCAAGAGTACGCAACGCTGAG 334





TITLE and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0078 row: M column: 05  
 Seg primer: CACACAGCAACACGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 503.

## FEATURES

Location/Qualifiers

1..503  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PMD42 (g14732114|gblaf129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 80 a 101 c 104 g 218 t  
 ORIGIN

## Alignment Scores:

Score: 0.152 Length: 503  
 Percent Similarity: 102.50 Matches: 46  
 Best Local Similarity: 46.678 Conservative: 38  
 Best Local Similarity: 25.568 Mismatches: 55  
 Query Match: 4.354 Indels: 41  
 DB: 12 Gaps: 10

US-09-807-459-2 (1-458) x A2344512 (1-503)

OY 223 ArgLeuGlnHisIleThrsSerSerTyRMetAspTyRcGlnIleProAlaLeu 242  
 DB 472 AAATGTGAGAACTCTATCAAGAAACAGACACTCATCATCAAGACACACATT 413  
 OY 243 ProLysPheAlaLysArgPheSerLeuMetValGlnArgLeuAlaThrValAla 262  
 DB 412 CCC-----ACTTCGCCCACTCA 395  
 OY 263 GlyTyValAspThrProTrpTyRysLysTyRMetLysLeuLysAsnPhMetVal 282  
 DB 394 GGGCAACCCCAACACACACC-----AAAAAGCTAGACCTGATTAAACATATCTCAG 341  
 OY 283 AsnArgValPheIleProThrLysLysPheAsn-----LysCluIleArgGlu 299  
 DB 340 ATGATGTGA---GAGACACCAAGAAAGACTTAATTAACCTCACTTAAGAAATACAGAG 284

OY 300 ProSer-----LysAlaLeuLysGluLysValSerThrAspThrLysAspLeu 315  
 DB 283 AACACTGCTAAACAGCTAGAACACATTAAGAAAGCAACAAATCCCTTAAGAAATTG 224  
 OY 316 PheGluAsnLysIleGlyngIlyThrValAspPheAsnLysGluIleArgAsp--- 334  
 DB 223 CAGGAAACACAAACCAACACAG-----GTGATGCAATGATTAACATCAACAGCTTA 170  
 OY 335 -----ProSerLysAlaLeuLysGluLysValSerAsnAlaLysAsp 349  
 DB 169 AAAGGGAAGTAGGCACAAATAAAGAAATCCAAAGAGGCAACGCTGATATAGAAAC 110  
 OY 350 LeuPheGluAsnLysIleGlyngIlyThrValAsp---PheIleAsnGluIleArg 368  
 DB 109 CTGAGCAAGAAA-----TCTGACACCATGATGCGACGATTCGACAAATGATCA 59  
 OY 369 AspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsn 388  
 DB 58 GAGATGGAAGAGGAAATC-----TCAGGTGCAAGAAATTTCAATAGAGAAC 14

RESULT 33 547 bp DNA linear GSS 24-JAN-2001  
 A2696548  
 A2696548/C  
 LOCUS  
 DEFINITION RPI-23-214L9, TJB RPI-23 Mus musculus genomic clone RPI-23-214L9,  
 DNA sequence.  
 ACCESSION A2696548  
 VERSION A2696548.1 GI:12411771  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 547)  
 REFERENCE Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akniet  
 and Fraser, C.M.  
 B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.  
 Mouse BAC End Sequences from Library RPI-23  
 Unpublished (1999)  
 Other GSSs: RPI-23-214L9, TVB  
 COMMENT Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@ligr.org

Clones are derived from the mouse BAC library RPI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (http://www.choi.org/bacpac/orderingframe.html). BAC end  
 page: http://www.tigr.org/cdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 214 row: L column: 9  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 Source Location/Qualifiers

1..547  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="RPI-23-214L9"  
 /clone\_lib="RPI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBAC3.6, site\_1;  
 EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methyase. Site  
 selected DNA was cloned into the pBAC3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 93 a 102 c 109 g 243 t  
 ORIGIN



```

Db 530 AAGAAGCACTTAAATATTGAAA-----ACA 559
QY 360 ValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValSer 379
Db 560 TTTGGATTATTCCTTAATAGTTTGTGATGAT-----CAAGTGTCTAAGAAATGTCTCT 610
QY 380 ThrGlyAlaGluAspLeuPheGluAsnLysIle 390
Db 611 AATACAGTTGAAACAAATTATAGTAAACAATATG 643

RESULT 35
BH070091/c 683 bp DNA linear GSS 18-JUL-2001
LOCUS RPCI-24-275B18.TV RPCI-24 Mus musculus genomic clone RPCI-24-275B18
DEFINITION , DNA sequence.
ACCESSION BH070091
VERSION BH070091.1 GI:14889688
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 683)
Zhao,S., Nieman,W., Malek,J., Shatsman,S., Akimret,B., Levins,M.,
Tsegaye,G., Geer,K., Kroll,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other_GSSs: RPCI-24-275B18.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0200
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 275 row: B column: 18
Seq primer: T7
Class: BAC ends.

FEATURES
Location/Qualifiers
source 1..683
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-275B18"
/clone_1lb="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/notes="Vector: pTRABAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC library produced by Pieter de Jong. The
library was cloned in the pTRABAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 124 a 130 c 140 g 289 t
ORIGIN
Alignment Scores:
Pred. No.: 0.32 Length: 683
Score: 101.50 Matches: 61
Percent Similarity: 42.74% Conservative: 42
Best Local Similarity: 25.31% Mismatches: 73
Query Match: 4.30% Indels: 66
DB: 12 Gaps: 13

US-09-807-459-2 (1-458) x BH070091 (1-683)
QY 190 PheThrThrGlyLeuPheGlyTrpGlyIleLysArgAlaLeuLysGlnIleIleArgSer 209
:::||||| ||||| |||
:::

```

```

Db 602 TACACAGCTGTCTCCACAGA-----GTGCTAACACAG 570
QY 210 AsnLeuProLeuAspIleGlyThrGluHisSerValSerArgLeuGlnHisIleThrSer 229
Db 569 GCTAACACACTCAGACAGAGAACACAGCTCTGCGCAGACACAATTAACAATCTAACACAA 510
QY 230 SerTyrLysAspTyrMetAspThrGlnIleProAlaLeuProLysPheAlaLysArgPhe 249
Db 509 GCATTAATCCAGATGGCAAAAGCAAAATACAAATCTGCACACAGCAAACTAAGACT--- 453
QY 250 SerLeuMetValAlaGlnArgLeuAlaThrValAla-----GlyTyrValAspThr 267
Db 452 AACTTGGCATCATCAGAACCACTATTCACACAGAGCTGTCTGTGATGCCACACACA 393
QY 268 ProTyrTyrLysTyrTyrMetLysLeuLysAsnPheMetValAsnArgValPheIle 287
Db 392 -----CTAGAAAAGCAATATCTGATTTGAAATCATATCTCATGATGCAGTTA---GAG 342
QY 288 ProThrLysLysPhePheAsnLysGluIleArgGlu-----Pro 300
Db 341 GATTTTAAAGAGAGACATTAATTAATCATCTAAAGAAATACAGAGAACACAGCTAAAGAG 282
QY 301 SerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGlu-----Asn 318
Db 281 ATAGAACTCTCTTAAGAGAGAAACACAAAAATCCCTTAACAATTAACAAGAAAAAACACAC 222
QY 319 LysIleGlyGlnGlyThrValAspPhePhe-AsnLysGluIleArgAspProSerLysAl 338
Db 221 CAATATGTGTAAAGCA-----ATTAACCAAAACCATCAAGAT----- 185
QY 338 AlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnI 358
Db 184 -CTAAAATGCAAGTA-----GAAACATAGAGAAATACAAAGAGAGAGAC 141
QY 358 YThrValAspPhe-----IleAsnAs 365
Db 140 AACTCTGGAGATFAGAAACCTAGAAAGAAATACAGAGACCATAGATCGACATCAACCA 81
QY 365 ngIuIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLe 385
Db 80 CAGAAATTCAGAGAC-----ATGAGAGAGCAATCTCT-GGTCCAGCAAGATTC 37
QY 385 uphGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspPr 405
Db 36 CATAGAAAC-----ATGACACACAAACAATTAATAATCC 4
QY 405 o 405
Db 3 A 3

RESULT 36
A2667704/c 830 bp DNA linear GSS 14-DEC-2000
LOCUS ENTX09TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION A2667704
VERSION A2667704.1 GI:11804970
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 830)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HML:IMSS sheared DNA library
JOURNAL Unpublished (2000)
CONTACT: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org

```

Clones are derived from the *Entamoeba histolytica* HM1:IMSS sheared DNA library

Seq primer: M13-Forward  
Class: shotgun

High quality sequence start: 33  
High quality sequence stop: 815.

## FEATURES

SOURCE

Location/Qualifiers  
1..830

/organism="Entamoeba histolytica"  
/strain="HM1:IMSS"

/db\_xref="taxon:5759"

/clone\_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pROSL. Site 1. Bst I. Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell. Oxford University Press, 1993)."

BASE COUNT 298 a 85 c 100 g 347 t  
ORIGIN

## Alignment Scores:

Pred. No.: 0.434 Length: 830  
Score: 101.50 Matches: 76  
Percent Similarity: 36.53% Conservative: 46  
Best Local Similarity: 22.75% Mismatches: 105  
Query Match: 4.30% Indels: 107  
DB: 12 Gaps: 20

US-09-807-459-2 (1-458) x A2667704 (1-830)

QY 104 PheileleuphelysglSeraspAlaasnProalaasnSerthGlulysArpGheTrp 123  
Db 777 TTTGTTAAATTTAAAGAG-----AAAATTGTAATTTGACAGACATGTTT--- 730  
QY 124 MetarGpheaArgGlyLysasnHisSerlyrPhehIsAspLeuValPheasnLeu 143  
Db 729 -----ATTGTTTCTTTGACTTACGCT---GATCTCTCAA 700  
QY 144 GlulysasnValThrArgaspAlaaspAlaThrAspIleGlulysPhealSerTrp 163  
Db 699 TATAACAATATACACAATGCA-----ATGAGAGTATT---AAACAAATAC 655  
QY 164 LeuTrMetAlaThrLeu-----TyrTrlysrThrTrhAsnValaspGluPheGly 181  
Db 654 ATTTATTTTATTTCAATTCCAAATTTCAAAATATTATATATATTTACGAACATATG 595  
QY 182 Ala-----SerPhehIsnLysleuSerPhehTrhThcLylLeuPheGlyTrp 197  
Db 594 GCACCAACATGTTTAAGTTCTTTTCA-----GATCTCTCAA 568  
QY 198 GlyIlelArgAlaLeuLysGlnIlelelArgSerAsnLeuProLeuAspIleGlyThr 217  
Db 567 AGTATTCCTCGAGTTTATTTATTCATTCTTCACAAAATTA-----TTATCAGTT 517  
QY 218 GluHisSerValSerArgLeuGlnHisIleThrSerSerTrpLysAspTrpMetAspThr 237  
Db 516 CGATCAGAAATTTGAAGCTTATGTCAT-----CATATGGAACCA 478  
QY 238 GlnIleProAlaLeuProLysPheAlaLysArgPheSerLeuMetValGlnArgLeu 257  
Db 477 TCACAAGTAAATTTTATTTACATCCAAAAGAAATAT-----AAACCACTT 433  
QY 258 LeuAlaThrValAlaGlyTrValAspThrProTrpTrpLysLysTrpTrpMetLysLeu 277  
Db 432 ATACATGTGATGAGAGGATCTAGAT-----TATTTGTTATTTATATGACCAAAA 382

QY 278 LysasnPheMetValAsnArgValPheIleProTrpLysPhehAsnLysGluIle 297  
Db 381 AAGAAAT-----ATATATCCAAATTAATAATTTATTAATGCGCAATAT 340  
QY 298 ArgGluProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGlu 317  
Db 339 AAATTAACCAACATGAGTTT-----CCTATTGACTTTAAACCACTTTT--- 295  
QY 318 AsnLysIleGlyGlnGlyThrValaspPhehAsnLysGluIleArgAspProSerLys 337  
Db 294 -----ATGTTTGAATTTGAAGATTAACCAACCAAAA 262  
QY 338 AlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGln 357  
Db 261 GAGATTA-----GATAGTATTTATTTATGAGCAAGTTTAAACAAA 220  
QY 358 GlyThrValaspPheIleAsnAsn----- 365  
Db 219 AATAGTATGATGATTTCTTAAATTTATATGATTTGATTAATACATATGCTT 160  
QY 366 -----GluIleArgaspProSerLysAlaLeuIleArgLysValSerThrGly 381  
Db 159 CAAGAAGATCAAGAATATAACAGAT-----ATTATTTCTATATTTCA----- 118  
QY 382 AlaGluaspLeuPheGluAsnLysIleGlyGlnGlyThrVal-----AspPheIleAsn 399  
Db 117 -----TGTTTCAAAACATTTATTTACAAATTCATCAATCAATTTGAATTTATACA 67  
QY 400 AsnGluIleArgaspProSerLysAlaLeuIleArgLysVal 413  
Db 66 AATGAATTAATAATTAATATGATACATTAATTTATTAATTAATTA 25  
  
RESULT 37  
LOCUS A2891746/c 605 bp DNA linear GSS 05-MAR-2001  
DEFINITION RPCI-24-209J3.TV RPCI-24 Mus musculus genomic clone RPCI-24-209J3,  
A2891746  
ACCESSION A2891746  
VERSION A2891746.1 GI:13210691  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 605)  
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Atkin, B., Levins, M.,  
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
Russell, D., de Jong, P. and Fraser, C. M.  
Mouse BAC End Sequences from Library RPCI-24  
unpublished (1999)  
Other GSSs: RPCI-24-209J3.TJ  
CONTACT: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0200  
Email: szhac@tigr.org  
Library availability, please contact Pieter de Jong  
(pdejong@mail.cno.org). Clones may be purchased from BACPAC  
Resources (<http://www.chori.org/dacpac/orderingframe.htm>). BAC end  
plate: [http://mw.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://mw.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
1..605  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-209J3"

## FEATURES

SOURCE







**THIS PAGE BLANK (USPTO)**